

＜話す・書く＞力を測るための両者共通の 評価項目による評価方法

酒 井 峰 男

An Evaluation Way for Assessing Speaking and Writing Abilities
With Commonly Based Evaluation Items
SAKAI Mineo

Abstract

This paper shows the reflections, based on Relevance Theory and consumers' point of view, on two important elements such as Evaluation Items and Criteria for Grading Levels, to evaluate especially international students' speaking and writing proficiencies among Japanese language learners. Concerning the speaking, grounding on Relevance Theory and consumers' point of view, (1) we found out Evaluation Items to assess the speaking proficiency, and (2) we showed Criteria for Grading Levels based on these Evaluation Items. Concerning the writing, grounding on Evaluation Items of speaking, (3) we picked up Evaluation Items of writing, and (4) we proposed Criteria for Grading Levels of writing through Evaluation Items of writing. Consequently, this work could let us propose commonly based Evaluation Items of speaking and writing more coherently and comprehensively generalized way.

キーワード：プレースメント・テスト、話す・書く、関連性理論、評価項目、レベル判定基準

1. はじめに

本稿では、高等教育機関がプレースメント・テストによって留学生の日本語能力、「話す」「書く」を測る場合、具体的にどのようなプロセスを考慮して評価項目やレベル判定基準を考えていったらよいかということについて論じたい。ここでプレースメント・テストをあげた理由は、評価項目やレベル判定基準を考える場合、プレースメント・テストを考えるのが一番包括的で、特に全体のレベル判定の整合性を考慮する上で、プレースメント・テストほど適したものはないからである。プレースメント・テストについて考えるということは、少し大げさに言えば、あらゆるテストを射程に入れるということであり、到達目標や到達レベルといった各高等教育機関の教育政策の「理想像」が、プレースメント・テストに直接反映されていると言ってもいい。それはどんな学生を育てたいかという教育機関が掲げる理念とつながっているはずだからである。そういった意味合いを含めて、プレースメント・テストの中の「話す」「書く」の評価を考えていきたい。なお、「話す」のプレースメント・テストの実施方法としては、ここでは多人数を想定した一斉テスト⁽¹⁾を考える。個別方式、たとえば、インタビュー方式⁽²⁾は、40～50人以上⁽³⁾の多人

数に対しては実用性に欠けるので、ここでは考えない。

本稿では「話す」力、「書く」力の評価の方法を考えていくわけであるが、基本的な立場として、「話す」に対しても、「書く」に対しても、両者に共通した評価項目（評価の柱となるべきもので、評価する際の一つ一つの細目ではない。）を使って評価できると考える。その理由については第7章で詳しく述べるが、「話す」と「書く」は四技能（話す・書く・聞く・読む）の中でも、話し手、または、書き手が、何かについて発話したり、何かについて書いたりするアウトプット型の言語行為で、聞き手や読み手に対し、ある発話やあるテキストを通して自分の感情や意図や情報を相手に伝達するという点から見ると、両者とも全く同じカテゴリーに属する言語行為である。「話す」「書く」には、それぞれ同じように談話の結束性やそれを支える統語的構成が必要であり、また、語句や文や段落を通して談話の意味内容を伝えるという点で、両者の談話には良く似た共通点がある。「話す」力、「書く」力を評価する場合、基本的に同じ評価項目を使って評価できるのではないか。それを探るのが本研究の目的である。

そこで、その共通の評価項目を見出す方法についてであるが、まず、ある理論や考え方から評価項目を導き出し、次に、そこから各レベルの判定基準を設定したい。なお、評価項目とは、評価者は被験者の何に注意を払い、こういった概念に焦点を当てて評価するか、その概念のことである。評価における柱となるべきもので、評価する際の一つ一つの細目を意味するものではない。それをここでは評価項目と呼ぶ。たとえば、音声や文法の適切さとか、語彙の豊富さとかいった、評価の柱となるべき概念を評価項目とするということである。通常、評価項目は複数ある。

次に、レベル分けとレベル判定基準について説明する。ある大学に新しい留学生が100人来たと想定しよう。100人とも日本語運用能力が異なり、さらに全員が既習者で、留学先の大学で日本語を続けて勉強したいと思っているとしよう。受け入れ機関としては何をしなければならないか。分かり切ったことだが、彼らのレベルに合ったクラスを設定し、レベル分けのための試験を実施することである。どうしてレベル分けが必要なのか。それは、学生にとっては学習を、教員にとってはクラス運営を効果的に進めるためである。2012年4月現在、筆者が勤務する岡山大学では6つのレベルを設けている⁽⁴⁾。初級1、初級2、中級入門、中級1、中級2、上級である。超級はない。本稿ではこの6つのレベルに基づいた「話す」「書く」のレベル判定基準について述べていきたい。レベル判定基準とは、評価項目のそれぞれが、どのレベルではどういう形で現れるか、その記述である。評定尺度の記述である。順番としては、まず「話す」について述べ、次に「書く」について述べる。

2. 「関連性理論」と「消費者の立場から見た事物の評価」

「話す」力の評価法についてはOPI⁽⁵⁾が広く知られており、牧野（2001：8-24）に詳しい。日本ではプレースメント・テストにOPIが使われている⁽⁶⁾と言われている。一方、「書く」力の評価法については菊池（1991：319-20）の評価法、また、それを参考にした石田（1992：154-5）の評価法⁽⁷⁾、また、田中、長坂（2006：270）が参考になる。第4章で本稿の評価法とOPIの評価法とを比較したいと思っているが、筆者の疑問は、例えば評価項目について、「話す」なら「話す」の、「書く」なら「書く」のそれらの分析的評価を行うための各評価項目が、一体どういう理由に基づいて、それらの項目なり、項目数が決定されているのかということである。「話す」について言えば、なぜ話題⁽⁸⁾や正確性がそれほど重要な評価項目になるのか。なぜ話題が豊富ならば上級なのか。日本語母語話者でも、話していて話題が少ない人はいくらでもいるが、その場合、彼らの日本語の「話す」力が弱いとされるのか。また、評価項目の区分の仕方であるが、話題が

豊富であることと、それに関連する語彙が豊富であることは同じこととして同じ評価項目の中に入れられないか。話題は語彙なくしては展開しない。さらに、また、正確性ということで、音声的な面、統語的な面、語用論的な面の正確性を一緒に考えていいのか。流暢さについては音声のリズムや間の取り方と関連しているので、発音の項目に分類できないか。

一方、「書く」力を評価する場合であるが、試験の時、「話す」の場合のようにさまざまなジャンルの話題を短時間で取り扱うことはできず、必然的に一つのジャンルのある話題を扱うことになり、テーマを設定して、それについて文章を書かせるのが一般的である。そして、「書く」の評価項目については、その談話全体の構成や展開の適切性や、文法や表記の正確性を重視することが多い。前述した菊池（1991）の作文の配点は、100点満点のうち、趣旨10点、正確性40点、内容50点となっており、正確性が全体の40%を占めている。菊池（1991）では、外国人が日本語で作文を書く場合に必要能力は二つあり、一つは「日本語能力」、もう一つは「文章能力」としている。そして、正確さには「日本語能力」が反映し、内容には「文章能力」が反映する度合いが高いと述べている。ここで考えたいことが二点あり、一つは評価基準の50%が内容となっていることである。菊池（1991）の内容の評価についての説明の中に、内容の評価に当たっては、原則的に日本語能力の評価は含めないものとするあり、日本語の誤り等については見ず、作文の内容の魅力、おもしろさ、豊かさ、確かさ、妥当性、適切さを総合的に見るという説明がある。被験者の「書く」力の中で一体何を評価すべきか、ということが問題になっていると思う。作文の内容の魅力、妥当性、適切さは、本稿のように日本語能力のレベル設定を考える場合、どの程度重要なものなのであろうか。特に日本語教育におけるレベル判定をする場合は、被験者が日本語学習者であるということを、むしろ前面に押し出し、内容については一切考慮しないという立場を取ることもできる。極言すると、「話す」にしろ「書く」にしろ、たとえつまらない内容であっても、音声面での聞き取りやすさ、表記面での読みやすさ、語彙面や統語面での豊かさや結束性、談話の一貫性等において、分かりやすく豊かであれば、被験者の日本語能力は一応評価できるのではないだろうか。内容そのものを大幅に評価対象に入れてしまうと、自由部分が多くなって評価の信頼性や妥当性が下がってしまう恐れが出てくる。もう一つの点は、評価項目の中に、先に取り上げた「内容」以外に「表現意欲」や「積極性」といった主観的な評価に近い項目が含まれていることである。評価項目の基準が今一つ明確ではなく、しかも、評価項目がダブっている印象を受ける。例えば、「表現意欲」や「積極性」は、語彙面や統語面での豊かさや結束性の中で評価できる。また、積極的に漢字を使った場合は、漢字表記の規範性などの項目を立てておけば、これら「表現意欲」や「積極性」は必要のない項目かもしれない。

本稿では、これらの「話す」「書く」の諸々の事情を鑑みて、評価する際の理論や考え方の拠所をできるだけ明確に示した上で、そこから、まずは評価項目を抽出し、続けてレベル判定基準を探っていきたい。統一的な見方から一貫性のある評価方法を探ることによって、「話す」「書く」の双方の談話の評価に応用できる、汎用性の高い評価方法を考えてみたい。まず、「話す」「書く」の言語運用をどうとらえるか。学習者から繰り出されるアウトプットの何を評価すべきなのか。そもそも高等機関として、学習者の何を評価したいのか。それを正しく評価するためにはどのような評価項目を立てるべきなのか。さらに、被験者のアウトプットをどのように判定するのか、そのレベル判定基準はどのように記述するのか。ここでは、理論から応用へ、「送り手」中心の評価から「受け手」中心の評価へ、「話す」評価方法から「書く」評価方法へ、という一連の流れの中で、「話す」「書く」の両者に共通する評価項目による「話す」「書く」の評価を試みたい。

2-1 関連性理論

語用論の中に含まれる関連性理論は会話成立の大原則を述べたものである。原理は2つあり、第1原理は、「関連性の認知の原理 (Cognitive Principle of Relevance)」であり、第2原理は、「関連性の伝達の原理 (Communicative Principle of Relevance)」である。さらに、第2原理には「最適な関連性の見込み (Presumption of Optimal Relevance)」が併記されている。D. スペルベル & D. ウイルソン (1999: 318-20) によると、第1原理は、「人間の認知は関連性が最大になるようにできている」ということである。聞き手は文脈の中で自分にとって最も関連性のあるものを選択し、最も効果的な情報処理を目指すものであるという。さらに、「生物機構はコストと利益のバランスが良くなるように、即ち、効率性が高くなるように進化したのであろうということであろう。」と述べ、人間の認知の進化を「コスト」と「利益」のバランスの中で説明している。このように、まずは認知の大原則が経済的概念で説明され、指摘されていることに注目しよう。言葉の経済性について、町田 (2004: 162) は「機能主義」と呼ばれたマルチネの例を引きながら、「経済性とは、コトバのしくみは人間ができるだけ労力を使わなくてもすむように、効率的に出来上がっているという性質のことを言います。」と説明している。また、町田 (2004: 207) は、言語の語順がどのように決まるかという言語研究の大問題に対し、例を引いて説明した後、「このように、ある言語の語順を決める重要な要素として、マルチネの提唱した経済性の原理が働いているのではないかと推測されるわけで、経済性という性質は、コトバのしくみを決定する原理を解明する上でこれまで以上に注目されなければならない要因なのではないかと思います。(下線は筆者による)」と述べている。ここ評価においても、この経済性という性質は大きな役割を演じていることを、2-2の「消費者の立場から見た事物の評価」でも改めて述べたい。

次に、この関連性理論の第2原理について述べる。第2原理は、「すべての意図明示的伝達行為はそれ自身の最適な関連性の見込みを伝達する」ということである。つまり、話し手が何か意図して発話をするということは、聞き手の推論が可能な範囲で聞き手にとって最も関連性のあることを聞き手に伝達していることになるということである。D. スペルベル & D. ウイルソン (1999: 189) では、「主人がお客にどうぞめしあがってくださいと言ったら、それは彼が勧めているものは食べられるものであり、しかも、実際食べるだけの値打ちがあるものだということを自動的に示唆していることになる。」という例を引いて説明している。また、第2原理に併記されている「最適な関連性の見込み」とは、前述のD. スペルベル & D. ウイルソン (1999: 331) によれば、(a)「意図明示的刺激は受け手がそれを処理する労力に見合うだけの関連性がある」、(b)「意図明示的刺激は伝達者の労力と優先事項に合致する最も関連性のあるものである」、と説明されている。つまり、話し手は聞き手がより少ない労力で最適な関連性が得られるよう文脈を定めなければならない、話し手は「聞き手への情報処理労力を低くせよ」という話し手に課せられた義務を果たさなくてはならない、ということである。聞き手の情報処理労力が少なければ少ないほど、関連性の度合いは高く、聞き手の情報処理労力が多ければ多いほど、関連性の度合いは低いということになる。できるだけ聞き手に無駄なエネルギーを消費させるな、というこの第2原理は、第1原理と同様、言葉とその経済性に深く関係している。聞き手中心の理論である関連性理論は、人間の認知や情報処理に関して、言葉の経済性にうまく適合している理論であることが分かる。聞き手中心⁽⁹⁾のこの理論は、そのまま言語能力の測定に応用できるのではないだろうか。「話す」の評価の場面を考えてみよう。発話者は被験者であり、聞き手は試験官となる。よって、発話者は試験官の情報処理労力を低くするような話し方をすればいいわけである。相手の負担を低くするような話し方、つまり、聞き手のエネルギーを浪費させないような話し方をすればいいわけ

ある。逆に、試験官から見れば、上級話者というのは関連性の高い話し方ができるレベルであり、初級話者というのは関連性の低い話し方しかできないレベルであるということになる。

以上、「話す」の評価の際に、評価の中心的概念として関連性理論が有効であると考え理由を演繹的にこの理論を応用して述べた。また、話し手を書き手に、聞き手を読み手に置き換えることによって、基本的に「書く」においても、この関連性理論が有効であると考え。しかし、関連性理論からだけでは具体的な評価項目は取り出せない。よって、次に評価項目を探るために、一体、評価者は何に焦点を当てて被験者のアウトプットである発話内容に対し評価を下すのかということについて考えてみたい。その際、消費者の立場から見た事物の評価について考える。事物を前にして消費者は何を問題にし、それらをどのように評価し、お金を払って事物を受け入れているのかということについて、3つの例を引き合いに出しながら、帰納的に考えていきたい。

2-2 消費者の立場から見た事物の評価

この章で、消費者の立場から見た事物の評価の例を取り出したのは、「聞き手への情報処理労力を低くせよ」という、聞き手が主導権を握る関連性理論は、消費者が事物を評価する際の立場をそのまま説明しているからである。製品の購入の際、消費者はできるだけ正確な情報を基に、数々の製品の中からある製品の品定め（＝評価）をし、購入し、消費する。その判断の基準となるものが、すなわち、評価項目である。消費者はどんな評価の基準を掲げて事物を評価するのか。どういう事項が保証されていれば、消費者はその製品は購入に値する、と評価するのか。社会を形成する消費者として事物を評価する際、そこには基準やルールがあるはずである。

1982年に国際消費者機構（C I）により「消費者の8つの権利と消費者の5つの責務」⁽¹⁰⁾が提唱された。「消費者の8つの権利」とは、1）生活の基本的ニーズが保障される権利、2）安全である権利、3）知らされる権利、4）選ぶことができる権利、5）意見が反映される権利、6）保障が受けられる権利、7）消費者教育を受ける権利、8）健全な環境の中で働き生活する権利、である。また、「消費者の5つの責務」とは、1）批判的意識、2）自己主張と行動、3）社会的関心、4）環境への自覚、5）連帯、である。日本では2004年に、この「消費者の8つの権利」の方が「消費者基本法」として制定された。これら「消費者の権利と責務」の項目の中で、本稿の「消費者の立場から見た事物の評価」に直接関係する項目は、「消費者の8つの権利」のうちの1）と7）を除いた6つの項目である。すなわち、2）安全、3）情報、4）選択、5）公正さ、6）補償、8）環境の安全、である。また、「消費者の5つの責務」のうちの1）批判的意識、が直接関係する。これは「価格、質、サービスに対して敏感で問題意識を持つ消費者になる、という責任（この行の訳と下線は筆者による）」と説明されている。先の6つの項目と、この3つの下線部分の項目を並べ替えてグループ化すると、1）価格、2）選択、3）安全、環境の安全、質、4）情報、公正さ、補償、サービス、の4つに分けられる。まとめて、それぞれ1）価格、2）種類の数と量、3）質、4）サービス、と命名する。2）の「選択」を「種類の数と量」、としたのは、消費者の選択権を補償するためには商品の種類の適切な数と適切な量が保証されねばならないからである。メニューの種類の数と各メニューの量の保証、と言えば分かりやすいであろうか。消費者側から見れば、生産者側が消費者の立場を第1に考え、消費者の負担を軽くするような製品を提供してくれるのが理想である。消費者にとって、製品は、経済的でしかも安全である物でなければならない。消費者側から見て、購入に余分な経済的コストがかからず、品質の割に値段が安いと感じられ、安心して長く使用できる物が理想である。「安くて、いい物」⁽¹¹⁾という表現に凝縮されているように、こういった製品が消費者にとっては一番購入しやすいであろう。な

ぜなら、関連性理論を応用すれば、それは消費者にとって一番負担が小さいからである。「安くて、いい物」の次に来るべき選択としては、「安いが、物が悪い」か、または、「高いが、いい物」となるであろう。「高くて、しかも、物が悪い」は消費者に見向きもされないであろう。この4つ目は「安くて、いい物」からはほど遠く、したがって、消費者の評価は最低となる。消費者への負担が前者の3つと比べて大きすぎるからである。「聞き手への情報処理労力を低くせよ」の命題に反するからである。

次に、レストランでの食事を考えてみよう。コックがいくらいいものだと信じて作った料理でも、お金を払って食べる客がどう思うかが重要である。料金が高かったり、量が少なかったり、食の安全性が保障されなかったり、店のサービスが悪くては、最終的に客からいい評価がもらえない。同じように、いくら自動車会社がいい車だと思って売り出しても、価格、種類、性能（燃費、運転しやすさ、安全性、頑丈さ、等）、デザイン（形状、色）、購入後のアフターサービス等で客が評価しなければ、いい車ではないことになる。大学の講義も同じである。‘消費者’である学生が、‘生産者’である教員の講義について評価する場合、学生は、授業料は妥当か⁽¹²⁾、講義内容、講義数はどうか、補講は保障されているか、履修者数は適当か、試験内容は適正か、成績評価は妥当か、講義室の設備は整っているか、などを総合的に判断して評価を下す。岡山大学が教員の個人評価において、学生授業評価アンケートを実施し、それを利用している⁽¹³⁾のも、大学は、大学という‘サービス業’に対する、学生という消費者から見た現場の評価を必要としているからである。我々は日常的に人や物を評価し、また、評価されながら生活している。評価とは決して専門家だけのものではない。ここでは評価されるものを‘商品’⁽¹⁴⁾と呼び、これらの‘商品’を対象にして、消費者側から見て、どんな評価項目を掲げれば、これらの‘商品’が評価できるかを見ることにする。また、‘商品’の種類によって、一般的に評価項目の間でより重視するものと、あまり重視しないものがあるが、それは考慮しない。あくまで、共通項として、最低限必要な客観的な評価項目を探ることにする。よって、個人の趣味や好みや印象に属するような主観的評価、たとえば、料理の味、店員の感じの良さ、店のインテリアの美しさ、車のデザインや色の良さ、講義内容がためになる、ならない、等の点については取り上げないことにする。

以下、‘商品’とそれを評価する際の評価項目を表1にまとめる。

表1：‘商品’とそれら进行评估する際の消費者から見た4つの評価項目

商品	評価項目1	評価項目2	評価項目3	評価項目4
食事	価格の適切性	メニューの種類の豊富さ、その量的な適切さ	食材の品質、安全性、食べやすさ	店員の対応や店のサービスの良さ
車	価格の適切性	車種の豊富さ、車体の量的な適切さ	車の性能、安全性、運転のしやすさ	販売員の対応やアフターサービスの良さ
大学の講義	授業料の適切性	講義の種類の豊富さ、その量的な適切さ	教員、講義の質、成績評価の分かりやすさ	教員の対応や大学のサービスの良さ

これらの‘商品’に対して、これら4つの客観的評価項目があれば、各‘商品’は十分評価できると考える。評価項目1の1) 価格の適切性、というのは、不当な値段で、消費者の財産が脅かされてはならないということである。評価項目2の2) 種類の豊富さとその量的な適切さ、というのは、種類の豊富さとその量的な適切性の中で選択の自由がなくてはならないということである。評価項目3の3) 質、というのは、品質や性能がよく、使いやすく安心して長く使用でき、環境が守られ、生活の安全が保障されるということである。評価項目4の4) サービスの良さ、というのは、消費者が差別なく公正に取り扱われ、‘商品’に関する情報がきちんと消費者に伝達され、

問題が生じた場合のフォローがなされることである。以上の4つの評価項目を簡略して取り出せば、前述したように、1) 価格、2) 種類の数と量、3) 質、4) サービス、となる。石けんの購入からマンションの購入まで、一般的に言って、‘商品’の価格が高くなればなるほど、評価項目4) が重要になってくる。これらの評価項目が全てクリアできれば、消費者の満足度は非常に高く、消費者は‘商品’について「経済的でしかも安全である」という判断を下す。それらの‘商品’は、消費者にとって精神的負担、経済的負担が小さいものとなり、したがって購入価値があると判断されやすくなる。関連性理論において、話し手に課せられた「聞き手への情報処理労力を低くせよ」という命題は、そのまま、生産者にとっては消費者に対し、「消費者への精神的、経済的負担を小さくせよ」という命題となる。そして、「消費者への精神的、経済的負担を小さくせよ」という命題を解決するためには、前記した4つの評価項目をクリアできれば良いということである。これら4つの評価項目の全ては「消費者への精神的、経済的負担を小さくせよ」という命題に、また、「聞き手への情報処理労力を低くせよ」という関連性理論の命題に合致している。両者の命題に共通している点は、「相手への負担を軽くせよ」、「相手のエネルギーを無駄に使わせるな」ということである。

以上、筆者は、言葉の評価に当たって、関連性理論という言語理論に加えて、消費者の立場から見た事物の評価⁽¹⁵⁾の仕方を見ることによって、‘商品’の評価に必要な評価項目を見た。次に、言語能力の中の「話す」力を評価する際に、‘商品’に対するこれら4つの評価項目がどのように応用できるかについて述べてみたい。

3. 「話す」の評価に必要な4つの評価項目：

ここでは、2-2で抽出された‘商品’に対する4つの評価項目、1) 価格、2) 種類の数と量、3) 質、4) サービス、について、それらが音声を介した発話に対してどのように応用できるかを述べる。繰り返すが、話し手には聞き手に対して「聞き手への情報処理労力を低くせよ」を可能にするような発話が求められている。

3-1 評価項目1：「音声の聞き取りやすさと発話のスムーズ性」

‘商品’に対する4つの評価項目の1)「価格」は、「音声」の聞き取りやすさと発話のスムーズ性、に相当する。‘商品’がお金を介しているのと同様に、発話は音声を介している。消費者にとって価格が安いということは、求めやすい、ということである。聞き手にとって聞き取りやすいということは、無駄なエネルギーが要らない、理解に負担がかからない、ということである。このように、音声的に見て、聞き手に負担をかけないものであるかどうか、つまり、聞き取りやすい発話であるかどうかの評価項目になる。聞き手が何度も「えっ？」と聞き返しをしなければ文脈が分からなくなるような発話者の不明瞭な発音や、間違って発話してしまった場合の意味不明の発音や、語頭の無駄な繰り返しや、話し手が意図しない冗長性（沈黙を含む）は聞き手を疲れさせる。リズムのない流暢性に欠ける話し方は聞き手をいらだたせる。このように、話し手の音声上の不適切さが原因で聞き手の文脈の理解、流れ、予測がさまたげられるような発話は、聞き手にエネルギーを無駄に浪費させ、よって、こういった話し方は、聞き手から見れば、それは経済性に欠け、効率的ではない、聞き手の立場を考慮しない‘ポライトネス’に欠ける話し方と判断される。聞き手にいらだちを引き起こす。それは「聞き手への情報処理労力を低くせよ」の命題に反する話し方であるからである。

以上、まずは、音声の聞き取りやすさかどうか、発話のスムーズ性かどうか、がチェック事項

となり、評価の対象となる。

3-2 評価項目2：「語彙の豊富さとその結束性」

‘商品’に対する4つの評価項目の2)「種類の数と量」は、「談話の種類の豊富さと語彙の豊富さとその結束性」に相当する。ただし、談話の種類（挨拶、スピーチ、小話、等）の豊富さについては、試験問題自体が談話の種類の数をそのまま限定してしまう場合が多いので、ここでは談話の種類の豊富さは省き、談話の展開の中での語彙の豊かさと語彙上の結束性を考える。談話の展開に十分対応できるだけの関連語彙の異なり数と量が十分であるか、語彙上の結束性⁽¹⁶⁾が十分であるか、を見る。例えば、バイク事故について話す場合は、バイク、スピード、カーブ、スリップ、対向車、センターライン、転倒など、語彙上の結束性のある関連語彙の異なり数が量的に十分でなければ、いくら統語的手段を駆使しても、聞き手にイメージを醸し出すことはできないであろう。語彙の種類が多いことは聞き手に対する情報源となり、さらにイメージを形作って、談話全体を膨らませる最も重要なものの一つである。語彙の豊かさと話の内容の豊かさとは関係する。テキスト言語学の発展に貢献したM.A.K.ハリディ/ルカイヤ・ハサン（1997）は、テキスト⁽¹⁷⁾の重要な要素として結束性（Cohesion）を掲げた。結束性には語彙的結束性と文法的結束性があるが、ここでは語彙的結束性を考える。あるトピックについての関連語彙が、聞き手への情報源として、また、イメージを醸し出す要素として、結束性を保ちつつ十分に使われている発話であるかどうかチェック事項となり、評価の対象となる。数詞や話し言葉に特有な縮約形も、語彙レベルのものとしてここに含められる。

3-3 評価項目3：「統語的結束性」

‘商品’に対する4つの評価項目の3)「質」は「統語的結束性」に相当する。発話の質がよいというのは、聞き手の文脈の流れの予想を助け、スムーズな理解に役立つ、適切な統語的構造を持った発話である、ということである。聞き手にとって理解しやすいということは、安心して文脈の予測が立てられるということである。前述したM.A.K.ハリディ/ルカイヤ・ハサン（1997）によると、テキストの結束性を高める手段としてさまざまな統語的手段があり、具体的には、代名詞、指示詞などの同一指示や、代用、省略、連結などが関与しているという。発話が、談話構成に必要なこれらの豊かで適切な統語的手段を持ったものであるかどうかチェック事項となり、評価の対象となる。当然、それらが形作るテキストの型、すなわち、文が単文レベルであるか、複文レベルであるか、段落レベルであるかも、豊かな統語的手段を持った、まとまりのある談話であるかどうかの決め手となり、評価の対象となる。談話構築の面からの一貫性のある文の展開もここに含まれる。

ここで注意しておきたいのは、「質がよい」の中には、「ためになる」「素質がある」「ユニークである」など、話の内容についての評価は評価項目には含めないということである。話の内容や被験者の意見それ自体についての善し悪しは一切問わない。日本語学習者の「話す」力を見るための評価対象とはしない。あくまでも第2言語としての日本語の運用能力という観点からのみ評価することにする。

3-4 評価項目4：「談話の管理」

‘商品’に対する4つの評価項目の4)「サービス」は「談話の管理」ができているかどうか、ということである。ここでは特に対人関係の中での発話であるということが十分考慮されているか

どうか、各場面にふさわしい適切な対人配慮が言語的にできるかどうか、ポライトネスも含めて聞き手の立場が十分配慮された言語運用がなされているか、コミュニケーション上で何か問題が生じた場合の言語的対応は補償されているか、が問われる。前述の評価項目1～3は談話の構成的能力の評価となっているが、評価項目4は談話の中の特に語用論的能力の評価に関係している。学生が研究室に来て、完璧な発音で「先生、今、暇ですか。」と言ったとしよう。この発話は評価項目の1～3はクリアしている。発音、語彙、統語のいずれを見ても合格である。しかし、ここには待遇の見地から見て、つまり、社会言語学的見地から見て、欠けているものがあり、聞き手に対して失礼な言い方、すなわち、聞き手に負担を強いる発話となっている。ここでは運用の仕方にもちがいがある。また、この場面で、たとえ「先生、今、ちょっとよろしいでしょうか。」という表現を使ったとしても、もし、例えば先生が自分の研究室で来客に対応していた場合にこの発話をしたとすると、やはり、ここでは運用の仕方にもちがいがある発話、戦略的なものが欠如している発話、談話の管理ができていない発話、つまり、結果的に見て、相手に負担を強いる発話となる。また、沈黙が相手に不安をもたらすような場面を考えてみよう。そこで、もし発話者に談話管理の能力が十分あれば、リップサービスとしての余剰性、冗長性が発話の中に出てくるであろう。冗談やユーモアを使う方策もここに入れられる。冗談やユーモアの発話は、言語的知識として難しいのではなく、周囲の状況をよく把握した上での、その場面での運用のタイミングや適切さが難しいのである。このように談話管理は重要なものである。談話管理には、会話戦略（あいづち、聞き返し、ターン、フィラー、訂正、追加、相手へのカバー、余剰性、等）や社会言語学的観点から見た言葉（敬体、普通体、待遇表現、男言葉・女言葉、罵倒語、等）の運用場面における時期やタイミングが考慮された使い方であるか、また、発話が字義どおりであるか、または言外の効果をねらっているか、などといった発話内行為もここに含まれる。ここでは、こういった語用論的能力や戦略的能力を通した、談話全体から見た場合の談話管理ができていない発話であるかどうかチェック事項となり、評価の対象となる。

以上の4つの評価項目は、「話し手はいかに聞き手に無駄な負担をかけないで発話できるか」という観点から抽出されたものである。いずれも「聞き手への情報処理労力を低くせよ」の命題から見ている。

これらの評価項目は、学習者のレベルが初級か、中級か、上級かを見るための、あくまでも日本語学習者を対象としたレベル判定のためのものである。例えば、就職試験の面接やその筆記試験では、日本語運用能力のレベル判定よりも、学生の思考内容を見るためのものである場合が多いので、そういった場合には、3-3で不問にした「ためになる」「素質がある」「ユニークである」等の評価項目、つまり、一般に言われるところの「話の内容の良さ」そのものを評価項目の中心に据えればよいと考える。このことは7-3の「書く」の評価項目についても当てはまる。

4. 本稿の4つの評価項目とOPIの4つの評価要素⁽¹⁸⁾の比較

牧野ほか（2001:9）の定義によると「OPIとは、外国語学習者の会話のタスク達成能力を、一般的な能力基準を参照しながら対面のインタビュー方式で判定するテストである。」ということである。一般的な能力基準を参照しながら、というOPIの理論のよりどころは、牧野ほか（2001:19-20）によると、チョムスキーのパフォーマンスとコンピテンスを分けずに一つにした、パフォーマンス中心の言語理論の研究を一方に、また、第2言語習得理論の研究をもう一方にということで、それらを下敷きにして能力基準が作成されているとある。それらの言語理論から引き出された4つの要素、すなわち、(1)「機能・タスク」、(2)「場面/話題」、(3)「テキストの型」、(4)

「正確さ」、の4つが評価要素としてあげられている。「機能・タスク」とは、被験者のタスク能力を意味し、その総合的なタスク能力を「話す」力とし、それを具体的に支えるのが、(2)、(3)、(4)ということである。また、(4)の「正確さ」の中の下位分類として6つの項目があげられており、それらは、(a) 流暢さ、(b) 文法、(c) 語彙、(d) 発音、(e) 社会言語学的能力、(f) 言語運用能力、となっている。一方、「話す」の評価に必要な評価項目を取り出すための本稿の理論や考え方の拠り所としては、前述したように、①「関連性理論」、②「消費者の立場からの事物の評価」、の2つに大きく依存し、そこから「話す」の評価に必要な4つの評価項目を抽出した。評価項目、評価要素は多ければ多いほど良いと言うものではない。Council of Europe (2004: 206)によると、4つか5つを超えると評価者側の認知的負荷が過大になるそうで、心理的な上限は7つだそうである。

以下、表2では、本稿の4つの評価項目がOPIの4つの評価要素のどの部分と関連しているかについて、比較してみることにする。

表2：「話す」力を評価するための本稿の4つの評価項目とOPIの4つの評価要素

「話す」力	
本稿の4つの評価項目	OPIの4つの評価要素
	(1)「機能・タスク」=総合的なタスク
評価項目1： 「音声の聞き取りやすさと発話のスムーズ性」	(4)「正確性」の中の (a) 流暢さ (d) 発音
評価項目2： 「語彙の豊富さとその結束性」	(2)「場面/話題」の中の「話題」 (4)「正確性」の中の (c) 語彙
評価項目3： 「統語的結束性」	(3)「テキストの型」 (4)「正確性」の中の (b) 文法
評価項目4： 「談話の管理」	(2)「場面/話題」の中の「場面」 (4)「正確性」の中の (e) 社会言語学的能力 (f) 言語運用能力

OPIは外国語学習者の「話す」力の言語運用能力を測るためのものであり、1986年がテストの実施開始年である。有資格のテスターにより、さまざまなトピックについて一対一のインタビュー形式（一人10分から30分）でテストは進む。その対象者は留学生だけに限られているわけではない。OPIの判定レベルは4つあり、初級、中級、上級、超級である。超級を除いて、初級、中級、上級にはそれぞれ下・中・上の3つの下位区分がある。よって全部で10のレベルがある。

一方、本稿の対象は留学生に絞られており、全部で6レベルで超級はなく、しかも一斉テストを想定している。試験問題である談話の種類は最初から限られ、そのため話題領域が制限されているので、話題の種類の豊富さは考慮していない。考慮していないが、語彙に大きな重きを置いている。つまり、OPIの「話題」を本稿では「語彙」に、また、OPIの「場面」を本稿では「管理」に置き換えている。「正確性」については、本稿ではこの「正確性」という言葉は使わず、「聞き手への情報処理労力を低くせよ」の命題のもとで、聞き手のスムーズな理解に役立つ、適切な語彙的、統語的結束性を持った発話であるかどうかという面からのみ、「正確性」を見ている。以上、本稿の4つの評価項目とOPIの4つの評価要素の比較を試みた。

5. 本稿の4つの評価項目から見てくる「言語能力」とは？

評価項目を探る作業は、そもそも言語能力 language ability をどうとらえているかという大きな問題につながる。人はコミュニケーションをする場合、「話す」場合も「書く」場合も、送り手は受け手に対して何を第一に考えなければならないだろうか。送り手が受け手に負う義務とは何か。

そのルールとは何か。さらに、受け手は送り手からのアウトプットの、つまり、送り手の言語能力から表出されたものの中の、どの部分を取り出して、それを送り手の言語能力に裏付けされたものとして評価するのか。本稿では、ここまでこう言った評価の中心的概念を、「関連性理論」と「消費者の立場から見た事物の評価」という二つの面から観察してきた。ここでの対象となっているのは被験者の言語能力である。が、そもそも言語能力とは何か、の説明には触れてこなかった。言語能力をどう評価するかは、言語能力をどう見るかに基づいたものでなければならない。後者の見方がなければ、前者は存在しない。両者は深く関連している。そこで、本稿の4つの評価項目から、逆にたどって、そこから導き出せる言語能力の定義を考えたい。

本稿の4つの評価項目を通して見た言語能力の定義は、結果的にL.F.バックマン/ A.S.パーマー（2000:77-86）に記されている言語能力の一部の見方に近いものとなっている。それによると、発話にしろ、文およびテキストにしろ、つまり、「話す」にしろ、「書く」にしろ、言語能力は①言語知識 language knowledge と、②方略的能力 strategic competence（またはメタ認知的方略 metacognitive strategy）の2つに大きく分けられ、①の言語知識は、さらに、a) 構造的知識 functional knowledge（文法的知識 grammatical knowledge およびテキスト的知識 textual knowledge）と、b) 語用論的知識 pragmatic knowledge（社会言語学的知識 sociolinguistic knowledge および機能的知識 functional knowledge）に分けられるということである。一方、②の方略的能力（メタ認知的方略）は目標設定 goal setting、アセスメント assessment、計画（立案） planning、の3つに分けられるという。この②の方略は、個人の問題解決のための方略で、言語使用の場面だけではなく、すべての認知的活動に含まれていると仮定される、としている。

本稿の4つの評価項目の中の評価項目1～3はL.F.バックマン/ A.S.パーマー（2000）による①の言語知識の中の a) の構造的知識の中に分類されている文法的知識のうち、音声の評価項目1：「音声の聞き取りやすさと発話のスムーズ性」に入れ、語彙の評価項目2：「語彙の豊富さとその結束性」に入れ、文法を評価項目3：「統語的結束性」に入れ、さらに、この評価項目3に、文法的知識の中のテキスト的知識（結束性の知識、会話構造の知識）を加えたものとなっている。また、評価項目4の「談話の管理」には、①の言語知識の中の構造的知識に並列されて述べられている語用論的知識の中の社会言語学的知識や、機能的知識（発話内能力とも呼ばれている）によって実際に表出される際の、実際に言語が運用される際の、場面や時期の適切さが含まれている。さらには、②の方略的能力（メタ認知的方略）がここに含まれている。ただし、②の方略的能力（メタ認知的方略）については、それを言語能力の重要な部分として認めるにしろ、前述したように、これは一般的な問題解決のための一般事象に対する認知的な働きであるので、3－4で述べた、本稿のようなレベル判定のための試験では、あまり考慮する必要はないかもしれない。

以上、第5章では本稿の4つの評価項目とL.F.バックマン/ A.S.パーマー（2000）によって示されている言語能力との関係を見、本稿が言語能力そのものをどう捉えているかの説明を試みた。

6. 4つの評価項目を基にした各レベルのレベル判定基準

理論的には4つの評価項目を基にして6つのレベルを設定した場合、 $6 \times 4 = 24$ の判定基準が必要となる。これらの判定基準は一種の指針で、枠組みのようなものである。

以下、表3のレベル判定基準は、酒井（2003）を大幅に書き直したもの⁽¹⁹⁾である。

表3：「話す」力を評価するための4つの評価項目から見た各レベルの判定基準

「話す」の 評価 項目 レベル	1：「音声の聞き取りやすさと発話のスムーズ性」 ・聞き手に負担をかけない発音で、スムーズに発話できるか。	2：「語彙の豊富さとその結束性」 ・聞き手に情報とイメージを与える結束性のある関連語彙が発話できるか。	3：「統語的結束性」 ・聞き手のスムーズな理解を促すための談話の統語的結束性と豊かな統語的手段を持った発話かどうか。	4：「談話の管理」 ・聞き手との対人関係の中での発話であることを考慮した談話全体の管理ができるか。
上級	非常に聞き取りやすい。聞き取りに負担がかからない。発話はスムーズである。	①自分の身の回りのこと、②自分の専門分野、③日常生活全般について関連語彙が十分に出来る。	談話に結束性があり、段落の形成、接続節間の連携、名詞修飾、引用、等ができ、構文が豊かである。	対人関係を考慮した、会話ストラテジー、待遇表現(例：普通体、敬体、等)が適宜に使える。
中級2	全体的に聞き取りやすい。聞き取りにくい場合があっても、聞き手の負担は小さい。発話はスムーズである。	①②はいいが、③の語彙はやや制限される。全体的に語彙量はある。	全体的に結束性のある話し方ができる。名詞修飾節、副詞節が使える、構文の豊かさを感じさせる。	対人関係を考慮した会話ストラテジー、待遇表現等がほぼ適宜に使える。
中級1	聞き取りにくい場合があり、修復は可能。スムーズさに欠ける場合あり。	①②はまあまあ。③の語彙は弱い。全体的に語彙不足だが、最低限の語彙は出せ、発話は保てる。	部分的に結束性が認められるが、全体としては弱い。複文の種類は少なく単文レベルの発話が多い。	対人関係を考慮した会話ストラテジー、待遇表現等が部分的に使える。
中級入門	聞き取りにくい場合あり。聞き手に負担がかかる場合あり。発話の修復ができず、スムーズでない場合あり。	①②でもかなり制限される。③の語彙は弱いので、情報性は低い。発話はかなり不十分ながら保持できる。	部分的にしか結束性が認められない。全体として構文の豊かさは感じさせない。	対人関係を考慮した会話ストラテジー、待遇表現等が部分的に使えるが、かなり制限される。
初級2	聞き取りにくい場合がしばしばある。修復できない場合もあり、意味不明となる場合もある。	①も②も制限される。③の語彙は全体的に少ない。同じ単語を繰り返す。数字が不正確。発話量は少ない。	部分的にしか結束性が認められない。単文の羅列が多い。簡単な複文はできる。が、構文の豊かさはない。	対人関係を考慮した会話ストラテジー、待遇表現等が部分的にも表出はむずかしい。
初級1	聞き取りにくい場合がよくある。修復できず、ためらいが多く、意味不明の場合がしばしばある。	①も②も制限される。発話量が少ないので、異なり語彙の量も極端に少ない。	部分的にしか結束性が認められない。語句の羅列が多い。構文の豊かさはない。	対人関係を考慮した会話ストラテジー、待遇表現等の表出はほとんどできない。

これは評価項目から見たレベル設定のための評価尺度の記述である。あくまでもプレースメント・テストにおける大枠である。ここで一点確認しておきたい。本来、これら各評価項目の各レベル内における重みは一樣ではなく、また、同じ項目内においても、重みはレベルによって変わってくるということである。つまり、プレースメント・テストではなく、各クラスの中で同レベルの受講生を対象に本稿の評価項目を使って学習者の「話す」力を評価する場合、各評価項目の重みを変えて応用できるということである。例えば、初級クラスの場合では、評価項目1の中の「音声の聞き取りやすさ」は他の3つの評価項目に比べて最重要視されるであろうし、同じ項目の中の「発話のスムーズ性」は初級クラスよりも中上級クラスにおいて、より強く求められるもので

あろう。同じように、評価項目2の中の「語彙上の結束性」も初級クラスよりも中上級クラスにより求められるであろう。このように、本稿の評価項目は、プレースメント・テストだけではなく、個々のクラスでの「話す」力の評価にも柔軟に応用できるものであると考える。

以上、第3章では「話す」力を測るための評価項目を、第4章では本稿の4つの評価項目とOPIの4つの評価要素の比較を、第5章では評価項目から見た言語能力を考察し、第6章ではレベル判定基準を記述した。本来であれば、これに続いて、被験者の発話を引き出すための「話す」の試験問題は具体的にどんな理論や考え方から抽出された問題なのか、さらにそれらの問題一つ一つからアウトプットされる被験者の発話の特徴を記した、各レベルにおける評価判定基準の一覧表とは一体どんなものか、また、実際に一斉テストとして各問題がどのような順で被験者に課されなければならないか、といったような事もここで同時に説明され、提示される必要がある。が、紙面の関係上、それらについてはここでは割愛する。酒井（2003）を参照して欲しい。

次章では「話す」の評価項目をベースにして、「書く」の評価項目について考える。

7. 「話す」の4つの評価項目から導き出される「書く」の4つの評価項目

ここでも関連性理論を中心にして、読み手に対する負担の軽重の観点から、「書く」力を評価するための評価項目を考える。前述の菊池（1991）は「大筋としては、対面テストも作文も、評価の基本的な方法はそれほど変わらないともいえそうに思われる」と述べている。「話す」も「書く」も基本的に同じアウトプット型の言語運用であるので、「書く」の評価項目も「話す」と同じ評価項目が利用できるのではないだろうか。第1章で述べたように、ここで言う評価項目とは評価の際の柱になるもののことである。個々の細分化されたチェック事項を意味するものではない。「話す」も「書く」も同じ柱が、つまり、同じ評価項目が使えるのではないかと考える理由は以下の通りである。そもそも「話す」と「書く」の違いは不連続ではなく、互いに連続しており、談話の種類によって、「話す」を採用したり、「書く」を採用したりしているわけである。何かを伝える、という点からは、どちらもコミュニケーションを目指している。どちらの様式を使うか、音声を使うか文字を使うか、の違いはあっても、両者とも、例えば構造的知識を必要とする語彙や文法上の結束性に関しては、どちらも談話の形成のために必要である。さらに、両者とも語句や節や段落を通してそれぞれの談話の内容を伝えているわけである。談話全体の視点から、談話をいかにコントロールし、管理し、構築していくかという語用論的知識も、同じように両者に必要である。もちろん、「話す」では、面と向かった対人関係が中心になり、話し手の個人的な個々の特別な場面における経験や感情等が再構築され、それが前面に出ることが多いし、また、読み手の立場を想像しながらの「書く」においては、題目が中心になり、私的な手紙やメール以外は、極力、個人的なことは避け、一般的な事象が理知的、客観的に分析され、それが前面に出ることが特にアカデミックな分野では多い。「話す」と「書く」のそれぞれの談話が持つ主観、情意、情報、分析等を、「話す」では発話（＝音声）を通して聞き手の情意や想像性に訴えようとし、「書く」では文字化されたものを通して読み手の知性や理性に訴えようとする違いはある。この点だけを見る限り、それぞれのゴールが全く違うように見える。しかし、これらはいずれも話し手や書き手が相手に伝えるために構築すべき一つの「談話」とであるという面から見れば、「話す」も「書く」も、どちらもコミュニケーションを目指す、ある構造を持った「談話」として扱える⁽²⁰⁾。両者を「談話」の観点から見れば、共通した評価項目が見出せるはずであるし、別々の評価項目を用いなければならない理由が見つからない。いずれも、聞き手または読み手への情報処理労力を低くせよ、の命題を守らなくてはいけないのは同じである。これらの理由により、両者の談

話を評価する場合、基本的に両者に共通の評価項目を用いて評価することができると考える。もし、同じ尺度で、つまり、「話す」評価項目から「書く」の評価項目を絞り込むことによって、両者に共通した、一貫性のある同じ項目で両者の評価ができれば、統一的な視点から両者の評価を考慮することができるようになり、少なくともテストの有用性⁽²¹⁾が高まることが考えられる。「話す」「書く」の両者の談話構造に注目しながら、いかに汎用性のある共通の評価項目を見つけるかが、唯一ここでは問題にすべきことだと考えたい。

ここでは、「話す」の発話を聞き手の視点から見たのと同様、「書く」も同じように読み手の視点から見る。すなわち、いかに書き手は読み手に負担がかからないような形で文字化された資料を提供できるかという観点から、評価項目の設定を試みる。前述した、田中、長坂（2006：270）の「書く」についての評価基準の中の一つに「読み手」という項目がわざわざ独立して掲げられているのは非常に興味深い。

以下では、「話す」の4つの評価項目を基礎にして、「書く」の4つの評価項目を取り出したい。

7-1 評価項目1：「漢字かな交じり文の表記と筆記の読みやすさ」

「話す」評価項目1の「音声の聞き取りやすさと発話のスムーズ性」に相当するものは、「書く」では「漢字かな交じり文の表記と筆記の読みやすさ」である。つまり、書かれた談話が、漢字、ひらがな、カタカナの混成文であることである。また、その筆記が読みやすく規範的であることである。ここの評価項目1は、音声を文字に置き換えただけである。漢字圏、非漢字圏を問わず、書き手から発信された文字資料が、いかに規範にのっとった形の字面で書かれているかが「書く」力のレベルを見る第一歩である。上級者を自負している学習者でも、漢字が一つもない文章しか書けないのであれば、いくら文字がきれいでも、文法的に正しくても、文章作成の技巧にすぐれていても、現代日本語の上級とは見なされない。文章の中に漢字がなければ、今日での日本語の規範的書き方とは見なされない。読み手に、負担のない、読みやすい文体だとは見なされない。よって、ここでの第1の評価項目として、どの程度の読みやすさを持った、規範に沿った漢字かな交じり文の表記ができるかがチェック事項となり、評価の対象となる。

7-2 評価項目2：「語彙の豊富さとその結束性」

「話す」の評価項目2の「語彙の豊富さとその結束性」に相当するものは、「書く」でも全く同じである。評価項目は同じであり、変わらない。「話す」テストの中では、インタビュー形式のテストのようにトピックが簡単に変えられる場合もあるが、通常、「書く」テストでは、談話の種類（ジャンルと言われるもので、感想文、報告文、意見文、論文、等）の中の一つのテーマを与えられ、それについてある一定の時間内で書く場合が多いので、トピックの種類の豊富さ等は評価項目として掲げられない。「話す」と同様、ここでも「語彙の豊富さとその結束性」が評価項目となる。課された文章の種類や目的に応じて、読み手の情報源やイメージ喚起の要素となる一連の語彙が、量的に十分に、しかも互いに結束性を持って活用されているかどうかチェック事項となり、評価の対象となる。

7-3 評価項目3：「統語的結束性」

「話す」の評価項目3の「統語的結束性」に相当するものは、「書く」でも全く同じである。評価項目は同じであり、変わらない。読み手への負担を小さくするべく、スムーズな理解に役立つ統語的結束性を持った構造の談話であるかが問われる。談話の結束性（段落、接続節、引用節、名

詞修飾、コソア、等）や、一貫性のある筋の構成などがチェック事項となり、評価の対象とされる。「話す」と同様、「書く」についてもその内容自体についての良し悪しは一切問わない。もし、内容を問う試験を実施したい場合には、この評価項目3の中に、内容に関する事項を特別に追加すればよいと考える。ここでは、あくまでも留学生を対象とした、日本語運用能力のレベルを測るプレースメント・テストのためのものである。

7-4 評価項目4：「談話の管理」

「話す」の評価項目4の「談話の管理」に相当するものは、「書く」でも全く同じである。評価項目は同じであり、変わらない。談話全体が読み手に配慮した、分かりやすい、規範的な書き方に準じた構成やレイアウトになっているかどうか、が問われる。構成やレイアウトが適切な個所に相応の仕方ではなされているかが問われる。具体的には、さまざまな記号の書き方、句読法、段落の始め方、タイトルの付け方、章の付け方、引用文の扱い方、参考文献の書き方、書き言葉、文体の統一、等がここに相当し、それらの運用方法の適切さを見る。さまざまな言語知識がどのように運用されているか、という管理面、運用面に焦点が当てられている。このほか、「話す」の評価項目4で、冗談やユーモア類の使用が扱われたように、「書く」でも同じで、レトリック的な技巧が実際の場で効果的に運用されているかどうか、ここ評価項目4のチェック事項となり、評価の対象となる。「話す」と同様に、「書く」の評価項目1～3は、談話の言語知識面から見た能力評価となっているが、評価項目4は談話の管理面から見た能力評価となっている。いずれも「読み手への情報処理労力を低くせよ」の命題から見ている。

以上、4つの「話す」の評価項目を基に、「書く」力を評価するための4つの評価項目が同じように取り出せることを示した。

8. 「書く」の4つの評価項目を基にした各レベルの判定基準

「話す」と同様、4つの評価項目を基にして6つのレベルを設定した場合、理論的には $6 \times 4 = 24$ の判定基準が必要となる。この表は評価項目に基づいた各レベルの目安であり、方向性を指し示したものである。「話す」と同様、一種の指針で、枠組みのようなものである。以下、4つの「書く」の評価項目を基にして、各レベルの記述を行う。

表4：「書く」力を評価するための4つの評価項目から見た各レベルの判定基準

「書く」の 評価 項目 レベル	1：「漢字かな交じり文の表記と筆記の読みやすさ」 ・読み手に負担をかけない規範的な表記法で書けるか。 筆記が判別しやすいか。	2：「語彙の豊富さとその結束性」 ・読み手に情報とイメージを与えるための結束性のある関連語彙を十分に書いているか。	3：「統語的結束性」 ・読み手のスムーズな理解を促すための談話の統語的結束性と豊かな統語的手段を持った書き方か。	4：「談話の管理」 ・読み手を考慮した談話全体の管理（書き言葉、文体、ジャンル、等）が規範に準じ、効果的に使用されているか。
上級	読む側に負担がない。漢字、ひらがな、カタカナが規範的である。筆記自体も読みやすい。	①自分の身の回り、 ②自分の専門分野、 ③一般社会、についての関連語彙が結束性を持って十分出せる。	談話に統語的結束性があり、段落の形成、接続節間の連携、名詞修飾、引用等が十分使える。	レポート文等、各種談話のジャンルの、規範的な書き方に準じて書ける。文体、書き言葉、等も統一されている。

中級2	読む側に負担があまりない。全体に漢字が規範より少ないが問題はない。筆記自体も読みやすい。	①②はいいが、③の語彙はやや制限される。全体的に語彙量あり。関連語彙の結束性もある。	談話に統語的結束性があり、段落の形成、接続節間の連携、名詞修飾、引用等が使える。	規範的な談話の書き方が大体できる。感想文、意見文、報告文、意見文が書ける。文体の統一はあるが、語句に話し言葉が混じる。
中級1	字面について読む側に負担が出る個所がある。漢字がかなり少なく、その分、ひらがなが多く占める。筆記において字面の間違いがある。	①②はまあまあ。③の語彙は弱い。全体的に語彙不足だが、最低限の語彙は出せる。関連語彙の結束性はやや制限される。	部分的に統語的結束性が認められるが、全体としては弱い。段落の形成、接続節間の連携、名詞修飾、引用等が一部使える。	談話の規範的な書き方が少しできる。十分ではないが、感想文、意見文、報告文の形ができる。文体の混同がある。語句に話し言葉が混じる。
中級入門	読む側に負担が出る個所が多々ある。全体に、かなが多く占める。筆記において字面の間違いがある。	①②でもかなり制限される。③の語彙はかなり弱い。数字はできる。最低限の語彙は出せる。関連語彙の結束性は制限される。	部分的にしか統語的結束性が認められない。段落の形成、接続節間の連携、名詞修飾、引用等の使用に大きな制限がある。	規範的な談話の書き方が少しできるのみ。全体がデス・マス調である。普通体の使用は困難である。書き言葉の使用が難しい。
初級2	読む側に負担がかかる。漢字はごくわずかで、かなが全体を占める。筆記において判別できない字体が多くある。	①②でもかなり制限される。③の語彙はほぼ現れない。語彙量が少ないので、情報量が少ない。同じ単語の繰り返しが目立つ。	全体として統語的結束性があるとは言えない。段落の形成、接続節間の連携、名詞修飾、引用等の使用は難しい。	規範的な談話の書き方ができない。日記文レベルはできる。全体がデス・マス調である。普通体は困難である。書き言葉の使用が難しい。
初級1	読む側に負担がかかる。漢字はまれで、かなが全体を占める。筆記において判別できない字体が多くある。	①②でもかなり制限される。③の語彙はほぼ現れない。語彙量が非常に限られている。同じ単語の繰り返しが目立つ。	統語的結束性があるとは言えない。段落の形成、接続節間の連携、名詞修飾、引用等の使用はほとんどできない。	談話の規範的な書き方ができない。日記文レベルでも限られている。デス・マス調である。普通体は困難である。書き言葉の使用はできない。

なお、この判定基準は漢字圏、非漢字圏を考慮していない。よって、評価項目1に関しては、漢字圏の学習者が有利になる場合も考えられる。つまり、漢字圏の学生は他の評価項目は低くても、評価項目1に関しては比較的高い判定がされる場合が考えられる。誰でも、得意、不得意があるので、評価項目1～4において、すべて同じレベルの判定を受けるということは、初級以外では少ないと考えた方が妥当である。これは「話す」にも当てはまる。

以上、「書く」の4つの評価項目から見た各レベルの判定基準を示した。

9. 結論と今後の課題

以上、日本語学習者の「話す」力と「書く」力を測るためのプレースメント・テストに必要な2つのこと、すなわち「評価項目」と「レベル判定基準」を複数の理論と考え方を応用して、必要な評価項目を統一的に抽出した。こうして、「話す」から「書く」の評価項目が取り出せることを示し、「話す」も「書く」も、ともに談話として見ることによって、同じ理論から取り出した一貫性、整合性のある共通した評価項目で、両者の評価項目が記述できることを示した。

実際のプレースメントの判定に当たっては、ある学習者に対するレベル判定が「話す」と「書

く」で同じではなく離れてしまう場合、例えば「話す」は上級に、「書く」は中級に判定されるような場合はどうするか、というような問題もあるだろうし、「話す」「書く」のそれぞれにおいて、例えば、問題1は中級1とか、問題2は中級2とかに分かれてしまう場合などもあるであろう。「話す」「書く」のそれぞれの採点尺度をどうするかという問題も生じる。問題が生じた場合、最終的に被験者をどのレベルに決めるかは、「話す」「書く」以外の他の技能の成績とも付け合わせ、場合によっては、個々のデータをもとに当事者と相談しながら決めるのが理想であろう。学内ではそれはまるっきり不可能なことではない。L.F.バックマン / A.S.パーマー（2000：15）が「言語テストに対する私たちの根本的な考え方」の中で述べているように、言語テストで大事なことは、「1. 言語テストを言語教育および言語使用と関連づける」であり、「6. テストの得点に基づいた決定はジレンマに満ちており、これらの問題に普遍的な答えなどないことを認識させる」とある。筆者の解釈によれば、前者は、言語テストはあくまで被験者の今後の言語教育に役立つものでなくてはならない、その反対であってはならない、ということであり、後者は、評価結果は唯一絶対的なものではない、ということである。一言で言えば、言語テストは被験者の言語学習意欲を削ぐようなものであってはならない、ということであろう。

現在、評価のスタンダード化⁽²²⁾が議論されている。今後、日本語学習者の数が増えるとともに、テストの種類も増え、日本語学習者の言語能力を測る必要性はますます高まると思われる。現在、教員と学習者が共有できる一貫性、透明性、整合性のある評価方法が、どの国のどの機関にとっても分かりやすい共通の尺度による評価方法が求められている。また、評価方法は誰にでもアクセスできるようなものが理想である。専門家だけのものであってはならない。本稿がそれに少しでも貢献できれば、幸いである。

最後に原稿の段階でさまざまな貴重な意見をくださった米国グアム大学人文社会科学学部の甲斐ますみ先生にこの場を借りてお礼申し上げたい。指摘を受けた点は多々あったが、十分に答えることができなかったのは、ひとえに筆者の力不足である。

[注]

- (1) 一斉テストとしては、LL教室でのテープ録音による方法を念頭に入れている。対話者1名と多数の被験者との同時一斉対話である。これについては酒井（2002）を参照のこと。今後は、設備が整えば、PCを使つての「話す」の一斉テストが考えられる。
- (2) インタビュー方式は、多人数を対象に実施するには実用的ではないと思われる。
- (3) 岡山大学では、毎学期、新しく入学してくる留学生が約40～50人いる。近い将来には100人になるかもしれない。どうやって「話す」のプレースメント・テストを実施するか、大きな問題である。筆者は、「話す」テストができない場合、筆記の「話すテスト＝談話構成テスト」（単文の連続から複文を構成するテスト）を実施しているが、もっといろいろな試みがあつていいと思う。
- (4) 岡山大学の6つのレベルの各授業は全学向け日本語補講として言語教育センターによって開講されている。履修時間数はどのレベルも半期約90時間（週4コマ/1コマ＝90分）である。その他、初級漢字1と初級漢字2が漢字クラスとして週1コマずつ、合わせて半期45時間分が開講されている。岡山大学の6つのレベルのそれぞれが終わった時点での日本語能力試験（N5～N1）との関係を大まかに述べると、初級2（約200時間の授業数）の到達目標はN5（150時間）、中級入門（約300時間の授業数）の到達目標はN4（300時間）、中級1（約400時

間の授業数)の到達目標はN3(450時間)への手前程度、中級2(約500時間の授業数)の到達目標はN3を越してN2(600時間)への道半ば、ということになるであろうか。一方、上級(約600時間の授業数)の到達目標はN2(600時間)ということになるであろう。岡山大学ではN1レベルまでの900時間という総時間数は用意していない。以上は、あくまで習得時間数を基にした大まかな説明である。

- (5) OPI (Oral Proficiency Interview) は総合的タスク能力試験であり、タスクが非常に重要視されている。タスクが全てである、と言ってもいいかもしれない。試験官は被験者にインタビューをしながら、被験者に課すタスクにより被験者のレベルを絞っていく。試験官にはOPIの資格が必要であり、逆に言えば、資格がないとOPIとしての試験を実施することはできない。資格(有効期限は4年)を得るには費用が20万円ほどかかり、期間も8カ月必要であると聞く。金銭面で余裕のない開発途上国の人は少数を除いて公式テスターになることは困難である。
- (6) このことは、嶋田(2010)のOPIの説明の中に書かれてある。
- (7) ただ、この作文採点基準表は週毎に行うような形成テストに対してのものである。
- (8) OPIがインタビュー方式を取っていることと、OPIが「話題」を重視していることには大きな関係がある。インタビュー方式のみが「話題」の自由な展開を可能にするからである。
- (9) 実際によくあることだが、話し手が分かりやすく説明できたと考えても、聞き手が理解できない場合がある。その際、聞き手は、「難しい」「理解できない」といったサインを出す場合がある。それを受けて、話し手側には言い直しが求められる。もし、言い直さなければ、対話は存在しなくなる。対話には聞き手の影響力が無視できない。
- (10) 日本語訳は日経部B P社 Webサイト: <http://www.nikkeibp.co.jp/sj/> を参照した。参考までに、「消費者の8つの権利」の内容1～8は英語で 1) The right to satisfaction of basic needs, 2) The right to safety, 3) The right to be informed, 4) The right to choose, 5) The right to be heard, 6) The right to redress, 7) The right to consumer education, 8) The right to a healthy environment, である。また、「消費者の5つの責務」の内容1～5は英語で 1) Critical Awareness, 2) Action, 3) Social Concern, 4) Environmental Awareness, 5) Solidarity, である。いずれも同サイトによる。
- (11) 「安くて、いい物」は理論ではなく一般常識の範疇に属するものである、よって、評価に関する考察にこういった一般常識に寄りかかるのは不適切だ、科学的ではない、と考える向きがあるかもしれない。しかし、ある理論を応用すれば、一般常識が分析できることもある。本稿の場合には、関連性理論により、「安くて、いい物」が、なぜ消費者の一般常識となっているのかが分析できる。結婚相手としての評価をする場合、一般的に男性の場合も女性の場合も、若くて、きれいで(ハンサムで)、やさしくて、お金に困っていない人、と答えるであろう。お年寄りで、不細工で、意地悪で、お金に困っている人、とは誰も答えないであろう。関連性理論を持ち出せば、この現象も説明できる。評価項目は誰にでも理解されるものでありたい。
- (12) 以前、岡大の法学部の学生で、大学の1回の講義に自分たち学生は一人あたりいくら大学に支払っているかを計算して、学内の刊行物に発表した学生がいた。確か1回の講義が二千元相当だったと思う。それ以後、学生が70人受講しているクラスで、筆者は毎回教室に入るたびに、自分は十四万円分の授業をしなければならないというプレッシャーを受けた。
- (13) 岡山大学は学生授業評価アンケートを教員の給料に一部反映させている。評価には責任が

伴うことを学生側にきちんと伝えておくことが必要である。

- (14) フロリアン・クルマス (1993: 96) は「言語は市場商品として、需要が増せば、市場性が良好となり、相場差益金が増えるというかぎり、他の商品とまったく同じ振る舞いをする。」というように、言語と商品の強い関係を述べている。事実、英語はすでに商品化されている。
- (15) 言葉と経済性の問題は特に新しいことではない。以下、町田 (2004: 163-165) を参考にしながら述べる。機能主義と呼ばれたマルチネ (1977) の有名な「二重分節」の学説は、言葉の経済性のしくみを前面に出したもので、二十数個の音素が何万、何十万の単語を作り、その単語が無限の文を生み出すという言葉のしくみを、音素—記号素 (第一次分節)、記号素—文 (第二次分節) とし、これを二重分節と命名して、言葉のしくみと経済性の問題について述べた。言葉と経済性の問題は、言語の運用の際の簡潔性や冗長性とも関連している。音声の縮約形をはじめ、指示や代用や省略など、簡潔性を助長し、冗長性を避けるための文法的手段は、さまざまな面で観察される。また、わざわざ冗長性を利用する場合も我々の言語運用にはある。エネルギーの無駄であり、エネルギーの浪費である冗長性と、その反対の簡潔性とは、言葉と経済性の問題に深く関係している。
- (16) 語彙上の結束性には、反復関係、反義関係、連語関係などが取り上げられている。
- (17) 本稿では、「テキスト」と「談話」は同じ意味合いで考えている。
- (18) OPIでは「評価要素」という言葉を使っている。
- (19) 酒井 (2003) では、評価項目として、(1)「情報の量」、(2)「情報の質」、(3)「伝達効率」、(4)「正確さ」、の4点を考えていた。(1)は語彙の異なり数、(2)は文法的結束性、(3)は伝達効率、(4)は音声・語彙・文法の正確性のほかに談話管理を入れていた。今回の評価項目との一番大きな違いは、伝達効率を取り下げ、談話のスムーズ性にしたことと、談話管理を独立させたことである。これについては、酒井 (2003) を参照のこと。
- (20) Alexandra Georgakopoulou & Dionysis Goutsos (1997: 36) は談話についてこう述べている。“In the light of numerous studies in various contexts, currently the standard view is that there are no absolute differences between spoken and written discourse. Spoken and written genres form a continuum and not the opposites of a rigid dichotomy.”
- (21) L.F. バックマン / A.S. パーマー (2000: 146-51) ではテストの有用性として、信頼性、構成概念妥当性、真正性、相互性、影響、実用性の6点を挙げている。
- (22) スタンダード化についてであるが、何かをスタンダード化ということは、何かをイデオロギー化することに繋がりやすく、そのため、言語テストを実施する側の社会的責任やその理念は常に問われ、検証されなければならない。また、それらに基づいて作成されるテスト自体の妥当性や、テストに関する理論的背景の説明も問われることになる。評価はその機関の理念や理想を表明するものである。その機関の目標を定めるものである。今後とも評価そのものについての議論が活発になることを期待したい。

[参考文献]

- 石田敏子 (1992) 『入門日本語テスト法』大修館書店
- 菊池康人 (1991) 「対面テストと作文の出題と採点 2. 作文」日本語教育学会編 (1991) 『日本語テストハンドブック』大修館書店 pp.305-324.
- 小池生夫主幹 (2003) 『応用言語学事典』研究社
- 国際交流基金 (2010) 『国際交流基金日本語教授法シリーズ8 書くことを教える』ひつじ書房

- 小林ミナ（2005）「学習者の日本語の評価」縫部義憲監修、水町伊佐男編集『講座・日本語教育学 第4巻 言語学習の支援』スリーエーネットワーク pp.152-170.
- 三枝令子（2005）「日本語・日本語教育に関する試験の現状と展望」縫部義憲監修、水島裕雅編集『講座・日本語教育学第1巻 文化の理解と言語の教育』スリーエーネットワーク pp.237-250.
- 酒井峰男（2011）『2011年8月15日～20日 岡山大学・フエ大学・外国語大学共催「日本語教育セミナー」実施報告書』
- 酒井峰男（2010）『日本語中級「話す」総合』岡山大学出版会
- 酒井峰男（2003）「全学日本語コースのレベル設定における諸問題」『岡山大学留学生センター紀要第10号』 pp.21-33.
- 酒井峰男（2002）「多人数にも応用できるディスコースを基盤にしたプロトタイプ方式による話す力の評価法」『岡山大学留学生センター紀要第9号』 pp.1-15.
- 嶋田和子（2010）「OPIについての説明」文化庁第35回国語分科会日本語教育小委員会・議事録
http://www.bunka.go.jp/kokugo_nihongo/bunkasingi/nihongo_34/gijiroku.html
- 谷口すみ子（2005）「言語能力」日本語教育学会編『新版日本語教育事典』大修館書店
- 田中真理、長坂朱美（2006）「第13章 第2言語としての日本語ライティング評価基準とその政策過程」『世界の言語テスト』 pp.251-276. くろしお出版
- 友松悦子（2008）『中級日本語学習者対象 小論文への12のステップ』スリーエーネットワーク
- 牧野成一、鎌田修、山内博之、齊藤真理子、萩原稚佳子、伊藤とく子、池崎美代子、中島和子（2001）『ACTFL-OPI入門』アルク
- 牧野成一（1986）『日本語教育』61号
- 鎌田修（1990）『日本語教育』71号
- 町田健（2004）『ソシユールと言語学 ことばはなぜ通じるのか』講談社現代新書
- D.スベルベル & D.ウイルソン（1999）『関連性理論—伝達と認知』研究社
- フロリアン・クルマス（1993）『ことばの経済学』大修館書店
- L.F.バックマン / A.S.パーマー（2000）『＜実践＞ 言語テスト作成法』大修館書店
- マルチネ, A.（1977）『共時言語学』白水社
- M.A.K.ハリディ/ルカイヤ・ハサン（1997）『テキストはどのように構成されるか』ひつじ書房
- Council of Europe/吉島茂、大橋理恵（訳者、編者）（2004）『外国語教育Ⅱ-外国語の学習、教授、評価のためのヨーロッパ共通参照枠-』朝日出版社
- Alexandra Georgakopoulou & Dionysis Goutsos（1997）Discourse Analysis. Edinburgh U.P.
- Council of Europe（2001）*Common European Framework of Reference for Languages : Learning teaching, assessment*. Cambridge University Press.